Illumination Invariant Face Recognition Using Thermal Infrared Imagery*

Diego A. Socolinsky† Lawrence B. Wolff‡ Joshua D. Neuheisel† Christopher K. Eveland‡

‡Equinox Corporation
9 West 57th Street
New York, NY 10019

†Equinox Corporation
207 East Redwood Street
Baltimore, MD 21202

{diego,wolff,jneuheisel,eveland}@equinoxsensors.com

Abstract

A key problem for face recognition has been accurate identification under variable illumination conditions. Conventional video cameras sense reflected light so that image grayvalues are a product of both intrinsic skin reflectivity and external incident illumination, thus obfuscating the intrinsic reflectivity of skin. Thermal emission from skin, on the other hand, is an intrinsic measurement that can be isolated from external illumination. We examine the invariance of Long-Wave InfraRed (LWIR) imagery with respect to different illumination conditions from the viewpoint of performance comparisons of two well-known face recognition algorithms applied to LWIR and visible imagery. We develop rigourous data collection protocols that formalize face recognition analysis for computer vision in the thermal IR.

1 Introduction

The potential for illumination invariant face recognition using thermal IR imagery has received little attention in the literature [1, 2]. The current paper quantifies such invariance by direct performance analysis and comparison of face recognition algorithms between visible and LWIR imagery.

It has often been noted in the literature [3, 2, 4] that variations in ambient illumination pose a significant challenge to existing face recognition algorithms. In fact, a variety of methods for compensating for such variations have been studied in order to boost recognition performance, including among others histogram equalization, laplacian transforms, gabor transforms and logaritmic transforms. All these techniques attempt to reduce the within-class variability introduced by changes in illumination, which severly degrades classification performance. Since thermal IR imagery is independent of ambient illumination, such problems do not exist.

To perform our experiments, we have developed a special Visible-IR sensor capable of taking simultaneous and co-registered images with both a visible CCD and a LWIR microbolometer. This is of particular significance for this test, since we are testing on exactly the same scenes for both the visible and IR recognition performance, not a boresighted pair of images.

In order to perform proper invariance analysis, it is necessary that thermal IR imagery be radiometrically calibrated. Radiometric calibration achieves a direct relationship between the grayvalue response at a pixel and the absolute amount of thermal emission from the corresponding scene element. This relationship is called responsivity. Thermal emission is measured as flux in units of power such as W/cm^2. The grayvalue response of thermal IR pixels for LWIR cameras is linear with respect to the amount of incident thermal radiation. The slope of this responsivity line is called the *gain* and the y-intercept is the *offset*. The gain and offset for each pixel on a thermal IR focal plane array is significantly variable across the array. That is, the linear relationship can be, and usually is, significantly different from pixel to pixel. This is illustrated in Figure 1 where both calibrated and uncalibrated images are shown of the same subject.

While radiometric calibration provides non-uniformity correction, the relationship back to a phisical property of the imaged object (its emmisivity) provides the further advantage of data where environmental factors contribute to a much lesser degree to within-class variability.

An added bonus of radiometric calibration for thermal IR is that it simplifies the problem of skin detection in cluttered scenes. The range of human body temperature is quite small, varying from $96°$F to $100°$F. We have found that skin temperature at $70°$F ambient room temperature to also have a small variable range from about $79°$F to $83°$F. Radiometric calibration makes it possible to perform an initial seg-

*This research was supported by the DARPA Human Identification at a Distance (HID) program, contract # DARPA/AFOSR F49620-01-C-0008.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2006		**00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Illumination Invariant Face Recognition Using Thermal Infrared Imagery	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Equinox Corporation,9 West 57th Street,New York,NY,10019**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**8**	

Figure 1: Calibrated (right) and uncalibrated LWIR images. There is significant pixel-wise difference in responsivity which is removed by the calibration process.

mentation of skin pixels in the correct temperature range.

2 Data Collection Procedure for Multi-Modal Imagery

The data used in the experiments performed for this paper was collected by the authors at the National Institute of Standards and Technology (NIST) during a two-day period. Visible and LWIR imagery was recorded with a prototype sensor developed by the authors, capable of imaging both modalities simultaneously through a common aperture. The output data consists of 240x320 pixel image pairs, co-registered to within $1/3$ pixel, where the visible image has 8 bits of grayscale resolution and the LWIR has 12 bits.

2.1 Calibration Procedures

All of the LWIR imagery was radiometrically calibrated. Since the responsivity of LWIR sensors are very linear, the pixelwise linear relation between grayvalues and flux can be computed by a process of two-point calibration. Images of a black-body radiator covering the entire field of view are taken at two known temperatures, and thus the gains and offsets are computed using the radiant flux for a black-body at a given temperature.

Note that this is only possible if the emissivity curve of a black-body as a function of temperature is known. This is given by Planck's Law, which states that the flux emmited at the wavelength λ by a blackbody at a given temperature T in $W/(cm^2 \mu m)$ is given by

$$W(\lambda, T) = \frac{2\pi hc^2}{\lambda^5 \left(e^{\frac{hc}{\lambda kT}} - 1 \right)} \tag{1}$$

where h is Planck's constant, k is Boltzman's constant, and c is the speed of light in a vacuum. To relate this to the flux

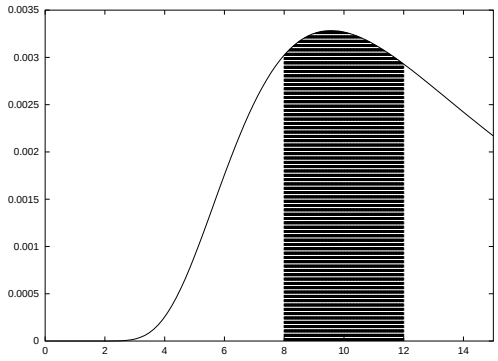

Figure 2: The Planck curve for a black-body at 303K (roughly skin temperature), with the area to be integrated for an 8-12μm sensor shaded.

observed by the sensor, the responsivity, $R(\lambda)$ of the sensor must be taken into account. This allows the flux observed by a specific sensor from a black-body at a given temperature to be determined:

$$W(T) = \int W(\lambda, T) R(\lambda) d\lambda \; . \tag{2}$$

For our sensor, the responsivity is very flat between 8 and 12 microns, so we can simply integrate Equation (1) for λ between 8 and 12. The Planck curve and the integration process are illustrated in Figure 2.

One can achieve marginally higher precision by taking measurements at multiple temperatures and obtaining the gains and offsets by least squares regression. For the case of thermal images of human faces, we take each of the two fixed temperatures to be below and above skin temperature, to obtain the highest quality calibration for skin levels of IR emmision.

It should be noted that a calibration has a limited life span. If a LWIR camera is radiometrically calibrated indoors, taking it outdoors where there is a significant ambient temperature difference will cause the gain and offset of linear responsivity of the focal plane array pixels to change. Therefore, radiometric calibration must be performed again. This effect is mostly due to the optics and FPA heating up, and causing the sensor to "see" more energy as a result. Also, suppose two separate data collections are taken with two separate LWIR cameras but with the exact same model number, identical camera settings and under the exact same environmental conditions. Nonetheless, no two thermal IR focal plane arrays are ever identical and the gain and offset of corresponding pixels between these separate cameras will be different. Yet another example; suppose two data collections are taken one year apart, with the same thermal IR camera. It is very likely that gain and offset character-

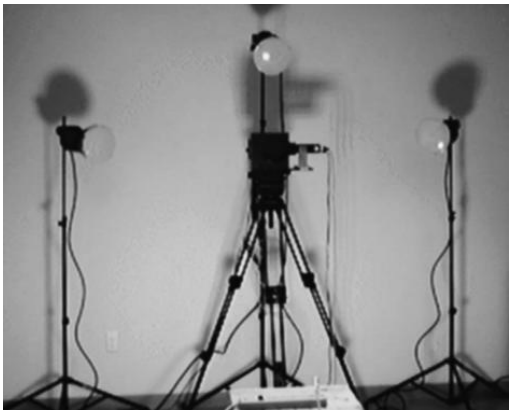

Figure 3: Camera and lighting setup for data collection.

Figure 4: Sample imagery from our data collection. Note that LWIR images are not radiometrically calibrated.

istics will have changed. Radiometric calibration standardizes all thermal IR data collections, whether they are taken under different environmental conditions or with different cameras or at different times.

The grayvalue for any thermal IR image is directly physically related to thermal emission flux which is a universal standard. This provides a standardized thermal IR biometric signature for humans. The images that face recognition algorithms can most benefit from in the thermal IR are not arrays of gray values, but rather arrays of corresponding thermal emission values. If there is no way to relate grayvalues to thermal emission values then it is not possible to do this.

2.2 The Collection Setup

For the collection of our images, we used the FBI mugshot standard light arrangement, shown in Figure 3. Image sequences were acquired with three illumination conditions: frontal, left lateral and right lateral. For each subject and illumination condition, a 40 frame, four second, image sequence was recorded while the subject pronounced the vowels looking towards the camera. After the initial 40 frames, three static shots were taken while the subject was asked to act out the expressions 'smile', 'frown', and 'surprise'. In addition, for those subjects who wore glasses, the entire process was done with and without glasses. Figure 4 shows a sampling of the data in both modalities.

A total of 115 subjects were imaged during a two-day period. After removing corrupted imagery from 24 subjects, our test database consists of over $25,000$ frames from 91 distinct subjects. Much of the data is highly correlated, so only specific portions of the database can be used for training and testing purposes without creating unrealistically simple recognition scenarios. This is explained in Section 4.

3 Algorithms Tested

Since the purpose of this paper is to remark on the viability of visible versus thermal IR imagery for face recognition, we used two standard algorithms for testing. We review them briefly in this section. Let $\{x_i\}_{i=1}^N$ be a set of N vectors in $\mathbb{R}^k$, for some fixed $k > 0$. Digital images are turned into vectors, converting a two-dimensional array into a one-dimensional one, by scanning in raster order.

Perhaps the most popular algorithm in the field is Eigenfaces [5]. Given a probe vector $p \in \mathbb{R}^k$ and a training set $\{t_i\}_{i=1}^N$, the Eigenfaces algorithm is simply a 1-nearest neighbor classifier with respect to the L^2 norm, where distances are computed between projections of the probe and training sets onto a fixed m-dimensional subspace $\mathcal{F} \subseteq \mathbb{R}^k$, known as the *face space*. The face space is computed by taking a (usually separate) set of training observations, and finding the unique ordered orthonormal basis of $\mathbb{R}^k$ that diagonalizes the covariance matrix of those observations, ordered by the variances along the corresponding one-dimensional subspaces. These vectors are known as *eigenfaces*. It is well-known that, for a fixed choice of n, the subspace spanned by the first n basis vectors is the one with lowest L^2 reconstruction error for any vector in the training set used to create the face space. Under the assumption that that training set is representative of all face images, the face space is taken to be a good low-dimensional approximation to the set of all possible face images under varying conditions.

We also performed tests using the ARENA algorithm [6]. ARENA is a simpler, appearance based algorithm which can also be characterized as a 1-nearest neighbor method. The algorithm proceeds by first reducing the dimensionality of training observations and probes alike. This is done by *pixelizing* the images to a very coarse resolution, replacing each pixel by the average gray value over a square neighbor-

hood. Once in the reduced-resolution space, of dimension n, 1-NN classification is performed with respect to the following semi-norm

$$L^0_\epsilon(x,y) = \sum_{j=1}^{n} 1_{[0,\epsilon]}\|x^j - y^j\|, \tag{3}$$

where 1_U denotes the indicator function of the set U.

4 Testing Procedure

In order to create interesting classification scenarios from our Visible/LWIR database, we constructed multiple query sets for testing and training. Frames 0, 3 and 9 (out of 40) from a given image sequence are referred to as vowel frames. Frames corresponding to 'smile', 'frown' and 'surprise' are referred to as expression frames. Our query criteria are as follows:

1. Vowel frames from all subjects, all illuminations.
2. Expression frames from all subjects, all illuminations.
3. Vowel frames from all subjects, frontal illumination.
4. Expression frames from all subjects, frontal illumination.
5. Vowel frames from all subjects, lateral illumination.
6. Expression frames from all subjects, lateral illumination.
7. Vowel frames from subjects wearing glasses, all illuminations.
8. Expression frames from subjects wearing glasses, all illuminations.
9. 500 random frames, arbitrary illumination.

The same queries were used to construct sets for visible and LWIR imagery, and all LWIR images were radiometrically calibrated. Locations of the eyes and the frenulum were semi-automatically located in all visible images, which also provided the corresponding locations in the co-registered LWIR frames. Using these feature locations, all images were geometrically transformed to a common standard, and cropped to eliminate all but the inner face. For the visible imagery, in addition to images processed as described above, we created a duplicate set to which we applied a multiscale version of center-surround processing [7] (a cousin of the Retinex algorithm [8]), to compensate for illumination variation.

The relation between vowel frames and expression frames is comparable to that between **fa** and **fb** sets in the FERET database, although our expression frames are often more different from vowel frames than in the FERET case. Frontal and lateral illumination frames are comparable to **fa** versus **fc** sets in FERET. Query set number 9 was used for face space computations for testing of the eigenfaces algorithm. Lastly, we should note that queries 7, 8 and 9 were only used as testing sets and not as training sets (except 7 versus 8), since the maximum possible correct classification performance achievable for those combinations is lower than 100%, and therefore those combinations were ignored to simplify the analysis.

All performance results reported below are for the top match. That is, a given probe is considered correctly classified if the closest image in the training set belongs to the same subject as the probe. Note that when using nearest-neighbor classifiers, one runs the risk that multiple training observations will be at the same distance from a probe. In particular it is possible to have multiple training observations at the minimum distance. This is especially likely when high-dimensional data is projected onto a low-dimensional space. In that case, it is possible to have false alarms even when considering only the top match, as it may not be unique. Let $\mathcal{T}$ be a training set and $\mathcal{P}$ a set of probes. For $p \in \mathcal{P}$, let m be the distance from p to the closest training observation, and $H_p = \{t \in \mathcal{T} \mid \mathrm{dist}(p,t) = m\}$. Define α_p to be 1 if any member of H belongs to the same class as p, and zero otherwise. Further define $\|H_p\|$ to be the number of distinct class labels among elements of H and $\#\mathcal{P}$ the number of probes in $\mathcal{P}$. With this notation, the correct classification rate and false alarm rate are given by

$$CC = \frac{1}{\#\mathcal{P}} \sum_{p\in\mathcal{P}} \alpha_p \ , \tag{4}$$

$$FA = \sum_{p\in\mathcal{P}} \frac{\|H_p\|}{\|H_p\| + \alpha_p} \tag{5}$$

5 Experimental results

For the ARENA algorithm, 240x320 images (in each modality) were pixelized and reduced to 15x20 pixels. We experimented with multiple values for the parameter ϵ, in the L^0_ϵ norm, and found that best performance was obtained with $\epsilon = 5$ for visible imagery and $\epsilon = 10$ for LWIR. This is consistent with the empirically observed fact that the LWIR imagery from this data set has an average effective dynamic-range of about 500 grayvalues, versus the 256 for the visible. Comparing Tables 1 and 2, we see that the ARENA algorithm on visible imagery benefits greatly from pre-processing with the center-surround method. The mean, and minimum classification performance over our queries on unprocessed visible imagery are 72% and 13%, respectively, with the minimum occurring when training is done

on lateral illuminated vowel frames and testing on frontal illuminated vowel frames. For center-surround processed visible imagery, the mean and minimum classification performance for ARENA are 93% and 76%, respectively, with the minimum occurring with training on frontal illuminated expression frames and testing on lateral illuminated vowel frames.

All performance results below are reported in tabular format, where each column of a Table corresponds to a training set and each row to a testing set. The numbering of the rows and columns matches that of the list at the start of Section 4. The diagonal entries are ommitted, since in all cases the classifiers achieve perfect performance when trained and tested on the same set. Also, certain sets are not used for training, since they do not contain images of all subjects, and therefore the maximum possible classification perfomance is strictly lower than 100%. Such sets are useful when testing the ability of a classifier to determine whether a given probe has a match in the training set at all, but we do not consider that problem in the current article.

Table 3 shows classification performance for ARENA on LWIR imagery. The mean and minimum performance are 99% and 97%, respectively. The minimum in this case occurs when we train on expression frames with frontal illumination and test on vowel frames for subjects wearing glasses. It is not surprising that the lowest performance would occur for probe sets where subjects are wearing glasses, since glass is opaque in the LWIR (see Figure 4). However it is surprising that the lowest performance is still quite high.

Despite the big performance boost that center-surround processing affords the ARENA algorithm on visible imagery, such processing is not suitable for use in combination with Eigenfaces. Our experiments show that it reduces performance in all situations. This is probably due to the fact that center-surround processing acts partially like a high-pass filter, removing the low-frequency components which are normally heavily represented among the first few principal eigenfaces. Since results were so poor on pre-processed visible imagery, we do not report specifics below. For our experiments we took the 240x320 images and subsampled them to obtain 768-dimensional feature vectors.

In Table 4, we see classification performance of Eigenfaces on visible imagery. The mean and minimum performance in this case are 78% and 32%, respectively. Minimum performance occurs for training on lateral illuminated vowel frames and testing on frontal illuminated expression frames, similar to the ARENA case. Performance of Eigenfaces on LWIR imagery is displayed in Table 5. Mean and minimum performance are 96% and 87%, respectively. Interestingly, the minimum occurs for the same training/testing combination as for the visible imagery. We can see by comparing Tables 4 and 5, that Eigenfaces on LWIR is uniformly superior to Eigenfaces on visible imagery. Classification performance is on average 17 percentage points higher for LWIR imagery, with the best scenario yielding an improvement of 54 percentage points over visible imagery, while never underperforming it.

Figures 5 and 6 show the first five eigenfaces for the visible and LWIR face spaces, respectively. The visible eigenfaces have a (by now) familiar look, containing mostly low-frequency information, and coding partly for variation in illumination. The corresponding LWIR eigenfaces do not have the 'usual' characteristics. In particular, we see that the LWIR eigenfaces have fewer low frequency components an many more high frequency ones. We verified this numerically by examining the modulus of the 2-dimensional Fourier transforms of the first eigenfaces in each modality, although it should be reasonably clear by simply looking at the images.

It is very interesting to look at the spectra of the eigenspace decompositions for the visible and LWIR face spaces. The corresponding normalized cumulative sums for the first 100 dimensions is shown in Figure 7. It is easy to see that the vast majority of the variance of the data distribution is contained in a lower dimensional subspace for the LWIR than for the visible imagery. For example, a 6-dimensional subspace is sufficient to capture over 95% of the variance of the LWIR data, whereas a 36-dimensional subspace is necessary to capture the same variance for the visible imagery. This fact alone may be responsible in large part for the higher performance of Eigenfaces on LWIR imagery than visible one. Indeed, since Eigenfaces is a nearest-neighbor classifier, it suffers from the standard 'curse of dimensionality' problems in pattern recognition. It has been recently noted in the literature that the notion of nearest neighbor becomes unstable, and eventually may be meaningless for high dimensional search spaces [9, 10]. In fact, for many classification problems, as the dimensionality of the feature space grows, the distance from a given probe to its nearest and farthest neighbors become indistinguishable, thus rendering the classifier unusable. It is not known whether face recognition falls in this (large) category of problems, but it can be safely assumed that data which has a lower intrinsic dimensionality is better suited for classification problems where the class conditional densities must be inferred from a limited training set. In this context, LWIR imagery of human faces may simply be 'better behaved' than visible imagery, and thus better suited for classification.

We further compared the performance of the two classifiers on both modalities, specifically for training/testing set pairs where each set had a different illumination condition. That is we looked at pairs where, for example, the training set had lateral illumination and the testing set had frontal illumination. Average and minimum classifications perfor-

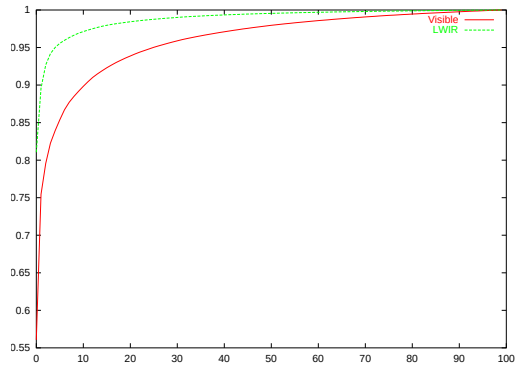

Figure 7: Normalized cumulative sums of the visible and LWIR eigenspectra.

mance are reported in Tables 6 and 7. Table 6 compares variation in illumination without major expression variation, while Table 7 corresponds to pairs where both illumination and expression are very different. We see that for both algorithms, LWIR imagery yields much higher classification performance over visible imagery. Indeed, for the variant illumination/expression experiment, LWIR Eigenfaces outperform visible Eigenfaces by more than 30 percentage points on average, and more than 50 for the worst-case scenario.

A combination of classifiers can often perform better than any one of its individual component classifiers. In fact there is a rich literature on combination of classifiers for identity verification, mostly geared towards combining voice and fingerprint, or voice and face biometrics (e.g.[11]). The main problem is how to combine the outputs of disparate classifier systems to produce a single decission. The ARENA face classifier is very well-suited to ensemble processing, since the L^0_ϵ norm is bounded above by the dimension of the space, therefore making interpoint distances from disparate distributions comparable to each other. To test this scenario, we constructed testing and training sets from the visible/LWIR image pairs. Now, an observation is not a single image, but a bi-modal pair. The testing set is composed of images where the subject is wearing glasses, and the training set contains images of the same subject but without glasses. This is a particularly challenging set, and the classification performance is 70.7% for visible ARENA, and 92.2% for LWIR ARENA. We can construct a distance between bi-modal observations simply as a weighted sum of the distances for visible and LWIR components, and then perform 1-NN classification on the new distance. By choosing the weighting parameter to be twice as large for the LWIR component as for the visible component, we can obtain a classification performance of 94.7%, an improvement of 2.5 percentage points over the LWIR classifier alone.

	1	2	3	4	5	6	7	8
1		0.986	0.557	0.514	0.741	0.717		
2	0.970		0.528	0.542	0.694	0.720		
3	1.000	0.988		0.988	0.226	0.182		
4	0.953	1.000	0.953		0.136	0.173		
5	1.000	0.986	0.334	0.276		0.986		
6	0.979	1.000	0.312	0.308	0.979			
7	1.000	0.990	0.604	0.532	0.733	0.729		0.969
8	0.992	1.000	0.552	0.564	0.692	0.716	0.985	
9	0.998	1.000	0.556	0.526	0.692	0.686		

Table 1: ARENA results on unprocessed visible imagery

	1	2	3	4	5	6	7	8
1		0.954	0.934	0.824	0.985	0.919		
2	0.944		0.830	0.914	0.903	0.970		
3	1.000	0.953		0.942	0.956	0.858		
4	0.946	1.000	0.922		0.833	0.913		
5	1.000	0.954	0.900	0.765		0.949		
6	0.943	1.000	0.783	0.870	0.938			
7	1.000	0.976	0.969	0.884	0.990	0.939		0.969
8	0.978	1.000	0.905	0.937	0.951	0.983	0.973	
9	0.995	0.961	0.904	0.858	0.971	0.935		

Table 2: ARENA results on center-surround processed visible imagery

	1	2	3	4	5	6	7	8
1		0.999	0.999	0.986	0.997	0.996		
2	0.997		0.993	0.987	0.995	0.996		
3	1.000	1.000		0.993	0.993	0.993		
4	1.000	1.000	1.000		0.993	0.990		
5	1.000	0.998	0.998	0.983		0.998		
6	0.996	1.000	0.990	0.980	0.996			
7	1.000	0.997	0.997	0.976	1.000	0.997		0.976
8	1.000	1.000	1.000	0.985	1.000	1.000	0.980	
9	1.000	0.998	0.996	0.978	1.000	0.998		

Table 3: ARENA results on LWIR imagery

	1	2	3	4	5	6	7	8
1		0.782	0.927	0.675	0.908	0.689		
2	0.701		0.598	0.909	0.601	0.867		
3	1.000	0.689		0.685	0.726	0.432		
4	0.593	1.000	0.591		0.319	0.607		
5	1.000	0.828	0.890	0.671		0.819		
6	0.756	1.000	0.602	0.863	0.745			
7	1.000	0.854	0.951	0.750	0.916	0.743		0.865
8	0.794	1.000	0.682	0.929	0.675	0.864	0.782	
9	0.920	0.856	0.846	0.760	0.826	0.768		

Table 4: Eigenfaces results on visible imagery

	1	2	3	4	5	6	7	8
1		0.973	0.981	0.922	0.995	0.962		
2	0.941		0.895	0.957	0.916	0.984		
3	1.000	0.974		0.942	0.986	0.949		
4	0.922	1.000	0.896		0.868	0.953		
5	1.000	0.973	0.972	0.912		0.969		
6	0.951	1.000	0.894	0.935	0.941			
7	1.000	0.976	0.997	0.967	1.000	0.972		0.956
8	0.956	1.000	0.929	0.983	0.936	0.987	0.966	
9	0.981	0.983	0.969	0.933	0.971	0.965		

Table 5: Eigenfaces results on LWIR imagery

Figure 5: First five visible eigenfaces.

Figure 6: First five LWIR eigenfaces.

	ARENA	Eigenfaces
Visible	0.874 / 0.783	0.663 / 0.432
LWIR	0.993 / 0.990	0.950 / 0.894

Table 6: Mean and minimum performance on experiments where the training and testing sets have different illumination but similar expressions

	ARENA	Eigenfaces
Visible	0.860 / 0.765	0.639 / 0.319
LWIR	0.990 / 0.980	0.933 / 0.868

Table 7: Mean and minimum performance on experiments where the training and testing sets have different illuminations and expressions

6 Conclusions

We presented a systematic performance analysis of two standard face recognition algorithms on visible and LWIR imagery. In support of our analysis, we performed a comprehensive data collection with a novel sensor system capable of acquiring co-registered visible/LWIR image pairs through a common aperture at video frame-rates. The data collection effort was designed to test the hypothesis that LWIR imagery would yield higher recognition performance under variable illumination conditions. Intra-personal variability was induced in the data by requiring that the subjects pronounce the vowels while being imaged, as well as having them act out severely variant facial expressions.

Dividing the data into multiple training and testing sets allowed us to gain some understanding of the shortcomings of each modality. As expected, variation in illumination conditions between training sets and probes resulted in markedly reduced performance for both classifiers on visible imagery. At the same time, such illumination variations have no significant effect on the performance on LWIR imagery. The presence or absence of glasses has more influence for LWIR than visible imagery, since glass is completely opaque in the LWIR. However, a variance-based classifier such as Eigenfaces can ignore their effect to a large extent by lowering the relevance of the area around the eyes within the face space.

Overall, classification performance on LWIR imagery appears to be superior to that on visible imagery, even for

testing/training pairs where there is no apparent reason for one to outperform the other. In the case of Eigenfaces, we can offer a plausible explanation for the superior performance in terms of the apparently lower intrinsic dimensionality of the data. Further experiments and data collections will be necessary to substantiate this conjecture, especially since the variance-based definition of 'intrinsic dimensionality' is a rather poor one. We intend to extend this investigation to the local dimensionality of LWIR face data as compared to its visible counterpart. Lower-dimensional face data would be a compelling reason for using LWIR imagery in face recognition systems. In essence, if we cannot beat the curse of dimensionality by statistical methods, we should perhaps be searching for alternative sources of lower-dimensional data with rich classification potential. LWIR face imagery may just be such a source.

It is possible that the high performance results we obtained are due to nature of our database. While our data may contain enough challenging cases for visible face classifiers, it may not cover the challenging situations for LWIR face classifiers. We intend to expand our collection and analysis effort in order to answer this question. One should note, however, that the data on which this study is based is very representative of common and important face recognition scenarios. For instance, indoor visitor identification and access to secure facilities and computer systems. The subjects we imaged were not prepared in any special way, and neither were the environmental conditions. Thus, though it may be possible to collect data which is much more challenging in the LWIR, it appears that this modality holds great promise under reasonable operating conditions. The main disadvantage of thermal-infrared-based biometric identification at the present time is the high price of the sensors. While the cost of thermal infrared sensors is significantly higher than that of visible ones, prices have been steadily declining over the last few years, and as volume increases they will continue to do so. This fact, in combination with their high performance, provides compelling reason for the deployment and continuing development of thermal biometric identification systems.

References

[1] F. J. Prokoski, "History, Current Status, and Future of Infrared Identification," in *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, Hilton Head, 2000.

[2] Joseph Wilder, P. Jonathon Phillips, Cunhong Jiang, and Stephen Wiener, "Comparison of Visible and Infra-Red Imagery for Face Recognition," in *Proceedings of 2nd International Conference on Automatic Face & Gesture Recognition*, Killington, VT, 1996, pp. 182–187.

[3] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," Tech. Rep. NISTIR 6264, National Instituite of Standards and Technology, 7 Jan. 1999.

[4] Yael Adini, Yael Moses, and Shimon Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 721–732, July 1997.

[5] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, pp. 71–86, 1991.

[6] Terence Sim, Rahul Sukthankar, Matthew D. Mullin, and Shumeet Baluja, "High-Performance Memory-based Face Recognition for Visitor Identification," in *Proceedings of IEEE Conf. Face and Gesture Recognition*, Grenoble, 2000.

[7] D. Fay et al. A. Waxman, A. Gove, "Color night vision: Opponent processing in the fusion of visible and IR imagery," *Neural Networks*, vol. 10, no. 1, pp. 1–6, 1997.

[8] Z. Rahman, D. Jobson, and G. Woodell, "Multiscale retinex for color rendition and dynamic range compression," in *SPIE Conference on Applications of Digital Image Processing XIX*, Denver, Nov. 1996.

[9] Allan Borodin, Rafail Otrovsky, and Yuval Rabani, "Lower Bounds for High Dimensional Nearest Neighbor Search and Related Problems," in *Proceedings of ACM Symposium on Theory of Computing*, 26 Apr. 1999, pp. 312–321.

[10] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft, "When is "Nearest Neighbor" Meaningful?," in *7th International Conference on Database Theory*, Jan. 1999.

[11] "Multi-Modal Person Authentication," in *Proceedings of Face Recognition: From Theory to Applications*, Stirling, 1997, NATO Advanced Study Institute.

[12] J. Prokoski, F., "Method and Apparatus for Recognizing and Classifying Individuals Based on Minutiae," 2001.